The Simple Repeat Poly(dT-dG) · Poly(dC-dA) Common to Eukaryotes Is Absent from Eubacteria and Archaebacteria and Rare in Protozoans

Julie Morris, Sidney R. Kushner, and Robert Ivarie
Department of Genetics, University of Georgia

Genomic DNA from a wide variety of prokaryotic and eukaryotic organisms has been assayed for the simple repeat sequence poly(dT-dG) · poly(dC-dA) by Southern blotting and DNA slot blot hybridizations. Consistent with findings of others, we have found the simple alternating sequence to be present in multiple copies in all organisms in the animal kingdom (e.g., mammals, reptiles, amphibians, fish, crustaceans, insects, jellyfish, nematodes). The TG element was also found in lower eukaryotes (Saccharomyces cerevisiae, Neurospora crassa, and Dictyostelium discoideum) and at a much lower frequency in protozoans (Oxytricha fallax and Tetrahymena thermophila). The sequence was also repeated in high copy number in a higher plant (Zea mays) as well as at very high levels in a unicellular green alga (Chlamydomonas reinhardtii). Although the copy number of the repeat per haploid genome was generally proportional to genome size, there was a greater-than-1,000-fold variation in the number of (TG)25/100-kb genomic DNA. By contrast, no eubacterium—including Myxococcus xanthus, whose life cycle is very similar to that of the slime mold Dictyostelium discoideum, and Halobacter volcanii, whose genome contains other repeated sequences—was found whose genomic DNA contained this sequence in detectable amounts. A computer search also failed to find the TG element in human mitochondrial DNA.

Introduction

Poly(dT-dG) · poly(dC-dA) is a Z DNA-forming sequence that has recently been found in multiple copies in the eukaryotic genome (Nishioka and Leder 1980; Slightom et al. 1980; Miesfeld et al. 1981; Shen et al. 1981; Hamada and Kakunaga 1982; Hamada et al. 1982; Rogers 1983; Sun et al. 1984). The first TG elements of substantial length were discovered in the 3' flanking region of a mouse kappa gene (Nishioka and Leder 1980) and in the nontranscribed spacer region of the mouse ribosomal RNA genes (Miesfeld et al. 1981). The element has since been shown to be present in the genomes of widely separated eukaryotes from mammals to yeast (Miesfeld et al. 1981; Hamada and Kakunaga 1982; Hamada et al. 1982). In its most common form, thymine and guanine residues alternate almost perfectly on one strand of the helix for 30–60 bp. In humans, the element is present at 50,000 copies/haploid genome interspersed throughout the genomic DNA (Hamada and Kakunaga 1982; Hamada et al. 1982; Sun et al. 1984). The element does not contain terminal repeats (Hamada et al. 1984a; Sun et al. 1984) and can reside in 5' or 3' flanking regions and in untranslated regions of genes but not in coding sequences (Nishioka and Leder 1980; Slightom et al. 1980;...

Its presence in “lower” eukaryotes such as yeast (Miesfeld et al. 1981; Hamada and Kakunaga 1982; Hamada et al. 1982) and slime mold (Miesfeld et al. 1981) as well as in “higher” eukaryotes suggests that it may serve a role(s) in the structure and function of the eukaryotic genome. Although its role is not understood, TG elements have been proposed to be a hot spot for recombination and gene conversion (Slightom et al. 1980; Shen et al. 1981); they undergo an enhanced level of homologous recombination in oversized SV40 DNA in CVI cells relative to comparable SV40 homologous repeats (Stringer 1985). The repeat element can also act as a transcriptional enhancer in a transient expression vector (Hamada et al. 1984b) for genes transcribed by RNA polymerase II. However, RNA polymerase III transcription was inhibited by the presence of a long element in regions flanking a tRNA pro gene from Caenorhabditis elegans (Santoro et al. 1984) and by a 9-bp TG,CG sequence beginning 20 bp upstream from a Xenopus laevis tRNA met gene (Hipskind and Clarkson 1983).

The observations of Hamada et al. (1982) suggested that the number of TG repeats in a particular organism was proportional to its haploid genome size. However, only a few organisms were analyzed and none were prokaryotes. In this study we have examined a large number of eukaryotic and prokaryotic organisms for the sequence and iteration frequency.

Material and Methods
Genomic DNA

High-molecular-weight genomic DNA from the organisms assayed in this report was generously supplied by several scientists: Methanococcus maripaludis DNA by B. Whitman, Myxococcus xanthus by L. Schimkets, Pseudomonas putida by M. Schell, Bacillus subtilis by B. Carlton, Drosophila melanogaster by J. McDonald, Anacystis nidulans and Chlamydomonas reinhardi by C. Hamilton, Neurospora crassa by J. Baum and N. Giles, Tetrahymena thermophila by C. Findly, Zea mays by J. Strommer, Caenorhabditis elegans by D. Hirsch and S. Karr at the University of Colorado, Mus musculus and Peromyscus maniculatus by S. Martin at the University of North Carolina, and Saccharomyces cerevisiae by A. Sugino at the National Institute of Environmental Health and Safety. B. Birmingham, N. Saunders, and J. Avise provided genomic DNA from Anguilla rostrata, Bagre marinus, Bufo terrestris, Trachemys scripta, Necturus alabamensis, Siren intermedia, and Limulus polyphemus.

Genomic DNA was extracted from other organisms as follows: Nuclei from axenic amoebae of Dictyostelium discoideum were isolated as described (Sogin and Olsen 1980), except that 0.2% Nonidet P-40 (Particle Data Laboratories, Ltd.) was used to lyse the cells; HeLa cell nuclei were isolated by low-speed centrifugation after lysis in 50 mM Tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl, pH 7.8), 0.1 M NaCl, 5 mM MgCl2, and 0.5% Nonidet P-40. Isolated nuclei or harvested sperm from Arbacia punctulata (Gulf Specimens, Inc.) were then lysed by the addition of sodium dodecyl sulfate (SDS) to 0.4% and digested with 20 µg/ml protease K at 55 C for 4–8 h. Escherichia coli were lysed by lysozyme–ethylenediamine tetraacetic acid treatment (Maniatis et al. 1982, p. 88), while Halobacter volcanii were lysed osmotically by gentle resuspension in 25 mM Tris-HCl (pH 7.8), 0.1 M NaCl, and 5 mM ethylenediamine tetraacetic acid. DNA was isolated from all lysates by extracting twice with redistilled phenol, twice with chloroform:isoamyl alcohol (24:1), and once with chloroform. DNA was precipitated with 70% ethanol and collected by low-speed cen-
trifugation or spooling. After dissolving in TE (10 mM Tris-HCl [pH 8.0] and 1 mM ethylenediamine tetraacetic acid), the DNA was digested with heat-treated RNase (20 μg/ml for 60 min at 37 C and then extracted once each with phenol and chloroform: isoamyl alcohol (24:1).

Southern Blots

DNAs were digested with various restriction enzymes, electrophoresed on 0.8% agarose gels in 40 mM Tris-acetate (pH 8.0), and transferred to nitrocellulose filters (Schleicher and Schuell) as described (Ivarie et al. 1983). One microgram of each DNA sample was electrophoresed in each lane. Filters were prehybridized for 16 h at 43 C in 5 x SSPE (1 x SSPE = 0.18M NaCl, 10 mM NaH2PO4 • H2O, 0.008% NaOH, and 1 mM ethylenediamine tetraacetic acid), 5 x Denhardt’s solution (1 mg/ml each polyvinylpyrrolidone [Sigma], Ficoll 400 [Sigma], and bovine serum albumin), 50% formamide, and 100 μg/ml yeast tRNA and then hybridized to [35S]-labeled poly(dT-dG)• poly(dC-dA) in the above solution for 24 h at 43 C. The labeled probe was prepared by nick translation using [35S]- (alpha-thio)-dATP as described (Ivarie et al. 1982; specific activity averaged 5 × 10⁸ cpm/μg). Blots were rinsed three times in 2 x SSPE and 0.1% SDS at 65 C. To assay DNA for TG elements at lower stringency, hybridization was performed in 40% formamide at 40 C and the last rinse of the filter was 2 x SSPE and 0.1% SDS at room temperature. Blots were visualized by autoradiography using Kodak XAR-5 X-ray film.

DNA Slot-Blot Hybridization

DNAs were denatured in 0.3 M NaOH for 60 min at 65 C, neutralized with an equal volume of 2 M ammonium acetate, and then spotted onto nitrocellulose filters using a Schleicher and Schuell minifold II slot blotton. Filters were probed as described above for Southern blots at high or low stringency as noted in the legend to each figure.

 Autoradiographic intensities of each DNA slot were measured at an optical density of 500 nm in a Beckman DI-8 spectrophotometer and compared to a known standard as described in Results. Blots were exposed to X-ray film for varying times to ensure that optical densities fell within the linear-response range of the film.

Results

Southern Blots

Genomic DNA from a variety of organisms was digested with EcoRI or PstI and assayed for TG elements by Southern blotting and hybridizing to [35S]-labeled poly(dT-dG)•poly(dC-dA). Typical results are illustrated in figure 1. Consistent with the findings of others (Miesfeld et al. 1981; Hamada and Kakunaga 1982; Hamada et al. 1982; Sun et al. 1984), human tumor cell DNA (HeLa) contained large amounts of the repeat sequence whereas Saccharomyces cerevisiae DNA contained fewer elements found in discrete restriction-enzyme bands. TG elements were also found in another fungus, Neurospora crassa, and at very high levels in the ciliated alga Chlamydomonas reinhardi. The element was also found in Zea mays (a flowering plant), Aurelia victoria (a jellyfish), and Drosophila melanogaster. Thus, by these limited criteria, the element appears to be ubiquitous in eukaryotic organisms.

By contrast, no detectable signal could be seen in the digests of total genomic DNA from the ciliated protozoan Tetrahymena thermophila or from three eubacteria (Escherichia coli, Bacillus subtilis and Myxococcus xanthus), two archaebacteria
(Methanococcus maripaludis and Halobacter volcanii), and the cyanobacterium Ana-
cystis nidulans (data not shown). The absence of signal from these organisms could
reflect either the absence of TG elements from their genomes or sequence degeneracy
such that they were not detected at the moderately high stringency conditions of
hybridization. In the case of Tetrahymena, the element might also be confined to the
micronucleus; >95% of total Tetrahymena DNA comes from the macronucleus, which
has lost 10%–15% of the sequence complexity of micronuclear DNA (Yao and Go-
rovsky 1974).

Southern Blots at Lower Stringency

To test these possibilities, micronuclear, macronuclear, and total Tetrahymena
DNA as well as bacterial DNAs were digested with HindIII and assayed for TG elements
by Southern blotting and hybridization at lower stringency (see Material and Methods). These conditions are routinely used to assay single-copy sequences in *Tetrahymena* DNA, given its low (25%) G + C content. The results are illustrated in figure 2. Discrete bands were seen in total *Tetrahymena* DNA after digestion with HindIII, while DNA from both the macro- and micronucleus gave positive signals even though neither was digested to completion. It also appears that micronuclear DNA contained substantially more TG elements than did macronuclear DNA.

It can also be seen in figure 2 that prokaryotic DNAs were largely devoid of detectable TG sequences; faint smears were visible in undigested DNA regions of the gel for all bacterial DNAs except *E. coli* DNA, which gave no signal. In this regard, bacteriophage lambda also gave weak signals under these conditions of hybridization (data not shown), and a computer-assisted sequence search found that the largest TG element in lambda DNA is (TG)$_4$, which occurs three times. In more than 150,000 nucleotides of *E. coli* DNA, only a small (TG)$_4$ was found by computer search. A similar result was found by scanning 11,730 bp of cyanobacterial DNA, while 11,700 bp of halobacterial DNA contained two short sequences, (CA)$_4$TTCA and

![Figure 2](image-url)

**Fig. 2.**—Southern blot hybridization of *Tetrahymena* and bacterial DNAs at lower stringency. Genomic DNA was digested with HindIII and assayed for TG elements by Southern hybridization as in fig. 1 but at lower stringency (see Methods). 1, *Tetrahymena* (total); 2, *Tetrahymena* micronucleus; 3, *Tetrahymena* macronucleus; 4, *Escherichia coli*; 5, *Methanococcus maripaludis*; 6, *Halobacter volcanii*; 7, *Pseudomonas putida*; 8, *Dictyostelium discoideum*.
CG(CA)$_3$CGCA, that might have been able to give a hybridization signal at low stringency. Although the exact size and sequence degeneracy that can give a positive signal under low stringency conditions is not known, a short CA-rich sequence in the rat prolactin gene consisting of (CA)$_n$GA[(CA)$_m$GA]$_4$ in intron C (Cooke and Baxter 1982; Maurer 1985) gave a positive signal comparable to that for bacterial DNAs under the same conditions of hybridization (J. Morris and I. Farrance, unpublished data). These observations suggest that prokaryotic DNA may contain very short TG runs that are assayable at low stringency.

Repetition Frequency of TG Elements

The conclusion that the copy number of the conserved TG element in eukaryotes was proportional to genome size was based on a small sample of eukaryotic DNAs (Hamada and Kakunaga 1982; Hamada et al. 1982). Accordingly, we have tested the general validity of the conclusion by assaying for the presence of the TG elements in 20 eukaryotes and measuring the copy number for 17 of those for which genome sizes were readily available (Yao et al. 1974; Fasman 1976; Lewin 1980). One test of the hypothesis involved small bacterial genomes, and, as already shown above, such organisms lacked detectable elements. The eukaryotes assayed included simple ones such as fungi and protozoans as well as complex metazoans in the plant and animal kingdoms. Genome sizes ranged from $10^7$ bp in Saccharomyces to $5 \times 10^9$ bp in the southern toad, Bufo terrestris.

Reiteration frequency was estimated by quantitative analysis of DNA slot blots of serial dilutions of the various DNAs (figs. 3, 4) and by comparison with a standard containing a known level of TG elements. A lambda-genomic clone of the rat prolactin gene (Cooke and Baxter 1982) was used as a standard; it contains two TG elements in its 5' flanking region (Maurer 1985). Allowing for 10% sequence mismatching, the recombinant phage contained 170 bp of assayable TG elements in 51,500 bp.

An example of a serial dilution of the standard is shown in figure 4 (col. 1). Slot blots of the standard DNA gave a linear optical-density curve proportional to time of exposure to X-ray film and to the amount of DNA blotted, as long as optical density did not exceed 0.35. Hence, optical-density values less than 0.35 were used in calculating the TG-element levels in the DNAs. Copy number was based on an average TG length of 50 bp in accordance with Hamada et al. (1982). As table 1 shows, copy numbers agreed well with published values (Hamada and Kakunaga 1982; Hamada et al. 1982): 48,000 versus 50,000 for human and HeLa cell DNA; $2 \times 10^5$ versus $10^5$ for mouse DNA; 500 versus 2,000 for Drosophila DNA; and 40 versus 100 for fungal DNA. Sun et al. (1984) estimated the frequency of TG elements as being $10^5$/haploid genome assuming a 40-bp length. It appears, therefore, that the rat prolactin gene provided a suitable standard for estimating genomic frequencies of this repeat family.

The TG-element frequency is plotted against genome size in figure 5. In general, reiteration frequency increased as the genome size increased. However, there was wide variation in the number of (TG)$_{25}$/100 kb of genomic DNA (table 1), ranging from 0.005 in protozoans to 25 in Chlamydomonas with a mean ± SE value of 3.9 ± 1.4. Even when the outlying values are ignored, the range was quite wide, being 0.1–13.2.

Species with large genomes and low ratios included the catfish Bagre marinus, the toad Bufo terrestris, and the flowering plant Zea mays. Furthermore, "lower" eukaryotes (e.g., fungi and protozoans) had ratios <0.15, yet the single-celled ciliate Chlamydomonas had the highest ratio at 25. Also, beyond a genome size of $10^5$ bp, there appeared to be a plateau in the frequency of TG repeats at $\sim 0.5-2 \times 10^5$/hap-
FIG. 3.—Dot-blot hybridizations of eukaryotic genomic DNAs. Known amounts of genomic DNA from various eukaryotes, as indicated on the side or top of each column, was blotted onto nitrocellulose and probed for TG elements as in fig. 1 at high stringency. Common names for most of the organisms listed below can be found in table 1. Panel A: 1, HeLa; 2, Chlamydomonas reinhardi; 3, Dictyostelium discoideum; 4, Neurospora crassa; 5, Tetrahymena thermophila (total DNA). Panel B: 1, Aurelia victoria; 2, Caenorhabditis elegans; 3, Drosophila melanogaster; 4, Limulus polyphemus; 5, Siren intermedia; 6, Necturus alabamensis; 7, Bufo terrestris; 8, Anguilla rostrata; 9, Bagre marinus; 10, Trachemys scripta; 11, Peromyscus maniculatus; 12, Mus domesticus.

loid genome. Whether the apparent plateau is in fact a true limit to the number of TG elements per genome needs to be tested by assaying organisms with genomes $>10^{10}$ bp.
FIG. 4.—Slot-blot hybridization of eukaryotic genomic DNA and the lambda-genomic rat prolactin-gene clone. Slot-blot hybridizations were carried out on the indicated amounts of DNA at high stringency as in fig. 3. 1, Rat prolactin gene (lambda rPRL); 2, HeLa; 3, Arbacia punctulata; 4, Neurospora crassa; 5, Oxytricha fallax macronucleus; 6, O. fallax micronucleus. Note that Oxytricha blots shown here came from gels exposed to film four times longer than the others.

TG Elements Are Rare in Ciliated Protozoan Genomes

Figure 5 also shows that the two ciliated protozoans, Oxytricha fallax and T. thermophila, have quite large genomes that contain small numbers of TG elements. Both protozoans are binucleate. The micronucleus functions in sexual reproduction and is transcriptionally inactive. It gives rise to the transcriptionally active macronucleus, which contains the bulk of the DNA of the vegetative cell by a process of DNA amplification and degradation in which the sequence complexity of macronuclear DNA is substantially reduced (Prescott et al. 1973; Yao and Gorovsky 1974). In Oxytricha (Prescott et al. 1973), >90% of the micronuclear sequences are lost, leaving DNA in gene-size 2.5-kb fragments, while in Tetrahymena (Yao and Gorovsky 1974) the sequence complexity is only reduced by ~15%. The data in table 1 and figure 5 imply that the reduction in sequence complexity in the macronucleus of Oxytricha also includes a comparable loss of TG elements. Hence, TG elements are unlikely to play a role in macronuclear sequence selection. Those that are in the macronucleus are probably retained by virtue of occurring in noncoding nucleotide sequences. Although the copy number has not been measured for Tetrahymena macro- and micronuclear DNA, a qualitatively similar conclusion can be drawn for Tetrahymena from the Southern blot analysis (fig. 2).

The observation that the two protozoans form a separate outlying series with respect to TG elements suggests a distinct line of descent of protozoans within the eukaryotic lineage. It is also noteworthy in this regard that ciliated protozoans also use an atypical genetic code in which the termination codons UAA and UAG specify glutamine (Caron and Meyer 1985; Horowitz and Gorovsky 1985; Preer et al. 1985), a finding that also suggests a very early divergence for this protozoan lineage.

Myxococcus and Halobacter Lack TG Elements

The bacterium M. xanthus and the cellular slime mold Dictyostelium discoideum have genome sizes within one order of magnitude of each other (5 × 10⁶ and 5.4 × 10⁷ bp, respectively) and remarkably similar life cycles (Bonner 1967, p. 43; Kaiser
Table 1
Genome Size of Various Organisms and Reiteration Frequency of TG Elements as Estimated by DNA Slot Blots

<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome Size(^a)</th>
<th>Total Copies(^b)</th>
<th>(TG)(_{25})/100 kb(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hela cells (human tumor)</td>
<td>3 \times 10^9</td>
<td>48,000</td>
<td>1.6</td>
</tr>
<tr>
<td>Mus domesticus (mouse)</td>
<td>2.7 \times 10^9</td>
<td>200,000</td>
<td>7.4</td>
</tr>
<tr>
<td>Bufo terrestris (toad)</td>
<td>5 \times 10^9</td>
<td>18,000</td>
<td>0.36</td>
</tr>
<tr>
<td>Anguilla rostrata (eel)</td>
<td>1.7 \times 10^9</td>
<td>60,000</td>
<td>3.5</td>
</tr>
<tr>
<td>Trachemys scripta (turtle)</td>
<td>2.3 \times 10^9</td>
<td>40,000(^d)</td>
<td>1.7</td>
</tr>
<tr>
<td>Bagre marinus (catfish)</td>
<td>2.2 \times 10^9</td>
<td>6,400</td>
<td>0.29</td>
</tr>
<tr>
<td>Limulus polyphemus (horseshoe crab)</td>
<td>4.6 \times 10^9</td>
<td>28,000</td>
<td>0.61</td>
</tr>
<tr>
<td>Arbacia punctulata (sea urchin)</td>
<td>7.2 \times 10^8</td>
<td>95,000</td>
<td>13.2</td>
</tr>
<tr>
<td>Aurelia victoria (jellyfish)</td>
<td>6.9 \times 10^8</td>
<td>515(^e)</td>
<td>0.08</td>
</tr>
<tr>
<td>Drosophila melanogaster (fruit fly)</td>
<td>1.5 \times 10^8</td>
<td>500</td>
<td>0.33</td>
</tr>
<tr>
<td>Caenorhabditis elegans (nematode)</td>
<td>8 \times 10^7</td>
<td>149</td>
<td>0.19</td>
</tr>
<tr>
<td>Dictyostelium discoideum (slime mold)</td>
<td>5.4 \times 10^7</td>
<td>115</td>
<td>0.21</td>
</tr>
<tr>
<td>Neurospora crassa (fungus)</td>
<td>2.7 \times 10^7</td>
<td>40</td>
<td>0.15</td>
</tr>
<tr>
<td>Oxytricha fallax (protozoan):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micronucleus</td>
<td>5 \times 10^8</td>
<td>25</td>
<td>0.005</td>
</tr>
<tr>
<td>Macronucleus</td>
<td>5 \times 10^7</td>
<td>6</td>
<td>0.012</td>
</tr>
<tr>
<td>Tetrahymena thermophila (protozoan)</td>
<td>2 \times 10^8</td>
<td>9(^f)</td>
<td>0.005</td>
</tr>
<tr>
<td>Chlamydomonas reinhardi (green alga)</td>
<td>1.2 \times 10^8</td>
<td>30,000</td>
<td>25</td>
</tr>
<tr>
<td>Zea mays (flowering plant)</td>
<td>5 \times 10^9</td>
<td>6,500</td>
<td>0.13</td>
</tr>
<tr>
<td>Bacteria (eu- and archaee-)</td>
<td>3.5 \times 10^6</td>
<td>(0.08)(^g)</td>
<td>...</td>
</tr>
<tr>
<td>Mitochondria (human)</td>
<td>1.6 \times 10^4</td>
<td>(0)(^h)</td>
<td>...</td>
</tr>
</tbody>
</table>

\(^a\) Genome size in base pairs came from chemical determinations and/or DNA reassociation kinetics (Fasman 1976, p. 284; Lewin 1980, p. 958).
\(^b\) Reiteration frequency of TG elements per haploid genome estimated from DNA slot blots assuming a length of 50 bp.
\(^c\) Ratio of (TG)\(_{25}\) per 100 kb of haploid genomic DNA.
\(^d\) Taken from chemically determined values for the green, wood, and snapping turtles of 5 pg/diploid nucleus (Fasman 1976, p. 284).
\(^e\) Value taken to be equivalent to that for another jellyfish, Aurelia aurita (Lewin 1980, p. 958).
\(^f\) Value taken as equivalent to that reported for Tetrahymena pyriformis (Yao et al. 1974).
\(^g\) Value taken as equivalent to that reported for Tetrahymena pyriformis (Yao et al. 1974).
\(^h\) A low, delectable signal was seen for 1 µg of DNA from Escherichia coli, Myxococcus xanthus, Halobacter volcanii, Pseudomonas putida, and Methanococcus maripaludis on long exposures, and these (differing by ±10%) have been averaged to estimate copy number for bacteria.

et al. 1979). Under appropriate environmental conditions both aggregate into multicellular fruiting bodies containing 100,000–1,000,000 cells, of which 10% eventually form spores. A difference is that Dictyostelium also forms a mobile slug that precedes fruiting-body formation. Despite these similarities, only Dictyostelium contains substantial levels of the TG element. The archaeabacterium Halobacter also has the eu-karyotic-like trait of containing repeated DNA sequences in its genome (Sapienza and Doolittle 1982), yet it does not contain detectable levels of the TG repeat.

**Discussion**
Although for the species analyzed here the copy number of TG repeats generally increased as genome size increased, no simple relationship between genome size and the number of (TG)\(_{25}\)'s was apparent. For example, the ratio of TG repeats/100 kb (excluding bacteria and mitochondria; table 1) varied from 0.005 to 25, or 5,000-fold. With two exceptions (Chlamydomonas reinhardi and Arbacia punctulata), invertebrates
had ratios <0.61; also with two exceptions (Bufo terrestris and Bagre marinus), vertebrates had ratios >1.6. The 5,000-fold difference in the ratio occurred between protozoans and Chlamydomonas that are very similar in that they are free-living, flagellated/ciliated, unicellular organisms. It appears, therefore, that the TG element is a dispensable part of the eukaryotic genome.

Nonetheless, the element was found in all eukaryotic genomes assayed. An unresolved problem concerns how the element initially became associated with the eukaryotic genome. One possibility involves the spontaneous nonreparable deamination of 5-methylcytosine to thymine. In higher eukaryotes, the only methylated base is 5-methylcytosine, which occurs most frequently in the dinucleotide CG (reviewed in Razin and Riggs 1980; Doerfler 1983). Several lines of evidence support the hypothesis that deamination of 5-methylcytosine to thymine is responsible for the scarcity of the
CG dinucleotide and for the overabundance of the TG and CA dinucleotides in eukaryotic genomes (Bird 1980; McClelland and Ivarie 1982). Hence, a heavily methylated alternating CG region of DNA might have gradually been converted to TGs in an ancestral organism. However, in extant organisms, no obvious correlation exists between an organism's content of 5-methylcytosine and TG elements. Drosophila and Neurospora, for instance, contain little if any 5-methylcytosine yet have fairly high levels of TG repeats in their genomes.

The absence of the repeat from prokaryotic genomes was not unexpected, given that most prokaryotic genomes lack reiterated sequences similar to those in eukaryotes (Sapienza and Doolittle 1982). Halobacter, however, is an exception because at least two species of the bacterium have been shown to contain several families of repeated sequences that are both clustered and dispersed in the genome (Sapienza and Doolittle 1982). Nonetheless, the simple TG repeat was undetectable in this prokaryote. Why the short repeat and, perhaps, others of the alternating purine/pyrimidine type are absent from prokaryotes is unknown. They may not be stable in circular chromosomes because of recombination- or replication-inducing deletions or deleterious rearrangements. It is known, for example, that some eukaryotic repetitive sequences are not stable when cloned in rec+ strains of Escherichia coli (Wyman et al. 1985). Furthermore, the Z DNA-forming sequence poly(dC-dG) is unstable in plasmids in E. coli when it is >30 bp in length (Klysik et al. 1982). Alternating C-G sequences in Z conformation have also been shown to block elongation of RNA transcripts when cloned downstream from a promoter. Poly(dT-dG) · poly(dC-dA), however, had no effect on transcription (Peck and Wang 1985). The foregoing assumes that TG repeats and other simple sequences arose occasionally during prokaryotic evolution and were subsequently selected against. It will be interesting to determine what happens when such elements are artificially introduced into mutant strains of E. coli defective in various steps of DNA metabolism.

Finally, the position of TG elements in gene families is not consistently conserved. Among six members of the human actin-gene family, three TG elements were in different introns of two genes and absent from the rest (Hamada et al. 1984a), whereas two nonallelic human gamma-globin genes harbor TG repeats in the same intron (Slightom et al. 1980). Thus, while TG elements may have arisen by descent in some gene families, their nonconservation in the actin family implies either that they were lost during subsequent evolution or that they arose de novo—and may continue to do so—by some other process(es) (i.e., 5-methylcytosine deamination or mistakes in DNA replication/repair/recombination).

Acknowledgments

The authors thank the many individuals who generously supplied genomic DNAs for this analysis and who are listed in the Material and Methods section; Norm Eberhardt and John Baxter for the lambda rat prolactin genomic-DNA clone; Marcus Fechheimer for Dictyostelium; and Iain Farrance for comments on the manuscript. This work was supported by grant CA34066 from the National Cancer Institute to R.I.

LITERATURE CITED


MIESFELD, R., M. KRYSZAL, and N. ARNHEIM. 1981. A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between human δ and β globin genes. Nucleic Acids Res. 9:5931-5947.


WES M. BROWN, reviewing editor

Received January 16, 1986; revision received March 13, 1986.